REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
	Technical Report	-

4. TITLE AND SUBTITLE
Network Evolution by Relevance and Importance Preferential Attachment

5a. CONTRACT NUMBER
W911NF-12-1-0546

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS
C Lim, Weituo Zhang

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Rensselaer Polytechnic Institute
110 8th Street

Troy, NY 12180 -3522

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
62449-NS.15

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

Relevance and importance are the two main factors when people find or build network connections. We propose an improved preferential attachment (PA) algorithm to take in consideration the relevance between vertices of the network measured by a given metric. We analyze the universal properties of the network class generalized by this algorithm and investigate two typical cases: scientific citation and between-city transportation. This is a brief report

15. SUBJECT TERMS
preferential attachment by importance and relevance

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Chjan Lim
UU	UU	UU	UU		19b. TELEPHONE NUMBER 518-276-6904

Report Title

Network Evolution by Relevance and Importance Preferential Attachment

ABSTRACT

Relevance and importance are the two main factors when people find or build network connections. We propose an improved preferential attachment (PA) algorithm to take in consideration the relevance between vertices of the network measured by a given metric. We analyze the universal properties of the network class generalized by this algorithm and investigate two typical cases: scientific citation and between-city transportation. This is a brief report of our research progress.

Network Evolution by Relevance and Importance Preferential Attachment

Weituo Zhang, Chjan Lim

August 6, 2014

Abstract

Relevance and importance are the two main factors when people find or build network connections. We propose an improved preferential attachment (PA) algorithm to take in consideration the relevance between vertices of the network measured by a given metric. We analyze the universal properties of the network class generalized by this algorithm and investigate two typical cases: scientific citation and between-city transportation. This is a brief report of our research progress.

1 Introduction

Relevance and Importance are the two main factors when people find or build network connections. One scenario is in the scientific research. For authors finding references, the importance of the articles and the relevance to their own issues should be both considered. Another scenario is in the decision making of constructing between-city transportation. We prefer to connect a city to other cities with higher connectivity but also want to reduce the expense by selecting nearby cities. In this paper, we propose an evolutionary network model with appealing properties that takes the both two factors into consideration. Our work is based on the "preferential attachment"(PA) algorithm invented by Barabasi, Albert. The classical preferential attachment starts with a network with N_0 vertices and m_0 edges. New vertex is successively added and attached to $m < m_0$ preexisting vertices. The probability of attaching to a vertex i is proportional to its degree k_i. This algorithm will naturally generate the network with power-law degree distribution $p(k) \sim k^{-\gamma}$ with $\gamma = 3$. There are many variations of the PA algorithm in the literature, and from which we conclude that the preferential attachment to high degree nodes, i.e. the "rich get richer" effect, is the essential reason for the emergence of scale free degree distribution. Besides, we suggest preferential attachment to relevant nodes, i.e. "connecting to things nearby" should be the reason that networks have clustering structures. Combining the both effects, it is hopeful to lead to network models with both scale free and high clustering properties, and it is the motivation of our work.

Although there is no rigorous definition of complex networks, many people consider the following three are the typical properties of complex networks: power-law degree distribution (scale free), high clustering coefficient (clustering), short average path-length (small world). A lot of efforts have been made to find network models which capture these properties. The following table summarizes the properties of several known network models.

Till now, not many network models satisfactorily capture all of the three typical properties. Some network models like Random Apollonian Network(RAN) do, but is totally

Network Model	scale free	clustering	small world
ER			$\surd$
BA	$\surd$		$\surd$
Lattice		$\surd$	
RGG		$\surd$	
SW		$\surd$	$\surd$
RAN	$\surd$	$\surd$	$\surd$

artificial without revealing the mechanism from which all the properties of the real world networks come. The RIPA model we proposed here have all of the three properties under certain conditions, and at the same time provides a natural reasoning of these properties. Further more, it also has a core-periphery structure which is an important feature of some real world networks like the world airline network (WAN).

In this paper, we will introduce our RIPA network model given by an evolution process, analyze several network properties, and compare this model with other network models and some empirical data.

2 Model

In this section we will describe the algorithm called Relevance and Importance Preferential Attachment(RIPA) which generate a class of complex networks. The RIPA, similar to the classical preferential attachment, starts with a initial network with N_0 vertices and m_0 edges. A new vertex is attached to m other vertices with the probability depending on the importance and relevance of those vertices.

In RIPA, the importance of a vertex is valued by its degree as in the classical preferential attachment. For the relevance, we introduce a metric space. In a metric space Ω, the distance between two elements $x, y \in \Omega$ is given by $d(x, y)$. Then their relevance $\rho(x, y)$ is defined as a non-increasing function of the distance between them $\rho(x, y) = f(d(x, y))$, satisfying $f(0) = 1$ and $f(\infty) = 0$. A typical example is $f(x) = e^{-x}$, but f can also have a power-law tail.

The centrality defined below measures the general influence of an element x on the whole space.

$$C(x) = \int_\Omega \rho(x, x')dx'$$

Centrality actually gives, in another sense, an "importance" according to the position in the underlining metric space instead of the connectivity to other vertices. In the scenario of the between-city transportation, centrality measures the physical geographical transportation condition of a position. In the scenario of scientific research, a research topic has high centrality means it is a bridge of many other fields. In this letter, we investigate some cases on metric spaces with constant centrality $C(x) \equiv C$. Examples are: (1)square with periodic boundary condition, (2) sphere in 3-d space, and (3) n-dimensional binary vector space with metric induced by L1 norm. In these spaces, there is no "center" position and every element is at an equivalent place.

A further restriction here for the relevance ρ and hence f is that the integral in the definition of centrality should be well-defined. This restriction is fairly important especially when we consider the large network limit.

In the RIPA, a new vertex j is attached to the preexisting vertex i by the probability

$$\Pi_{ij} = \frac{k_i \rho_{ij}}{z(x_j)}.$$

Here k_i is the degree of i indicating the importance and $\rho_{ij} = \rho(x_i, x_j)$ is the relevance between i, j. $z(x_j)$ is the normalization constant so that $\sum_i \Pi_{ij} = 1$. $z(x)$ is defined as a function on Ω called local partition by

$$z(x) = \sum_i k_i \rho(x_i, x).$$

The summation here goes over all existing vertices. A particular position $x \in \Omega$ with higher local partition $z(x)$ has more overall relevance to previous vertices, therefore may attract more interest. So we suggest $\mu(x)$, the probability of emergence of a new vertex at x, is proportional to $z(x)$

$$\mu(x) = \frac{z(x)}{Z},$$

where Z is the global partition function

$$Z = \int_\Omega z(x) dx = \int_\Omega \sum_j k_j \rho(x_j, x) dx = \sum_j k_j C(x_j).$$

We summarize the algorithm of RIPA as follows:

- 1. Begin with a network with N_0 nodes.
- 2. For $i = N_0 + 1$ to N

  2.1 Add a new node i at the position x with probability $\mu(x) = \frac{z(x)}{Z}$.

  2.2 Attach i to m preexisting nodes with probability $\Pi_{ij} = \frac{k_i \rho_{ij}}{z(x_j)}$.

In a metric space with constant centrality, we further have $Z = KC$ where $K = \sum_i k_i = m_0 + mt$ is the total number of degree in the network and grows linearly with time t. The expected change of the degree of the vertex i is given by

$$E\left[\frac{dk_i}{dt}\right] = \int_\Omega \Pi_{ij} \mu(x_j) dx_j = \int_\Omega \frac{k_i \rho_{ij}}{z(x_j)} \frac{z(x_j)}{Z} dx_j = \frac{k_i C(x_i)}{Z}.$$

The above equation shows that the degree of a vertex grows at a expected speed proportional to the current degree which is exactly the relation we have in standard preferential attachment algorithm. So we also obtain the power-law degree distribution $p(k) \sim k^{-\gamma}$ with $\gamma = 3$. Besides, the change of the local partition $z(x)$ comes from two parts: the growth of degrees of the existing vertices and the new vertex. When the centrality is constant C, we have

$$\begin{aligned} E\left[\frac{\partial z(x)}{\partial t}\right] &= \sum_i E\left[\frac{dk_i}{dt}\right] \rho(x_i, x) + m \int_\Omega \rho(x', x) \mu(x') dx' \\ &= \frac{C}{Z} \left(z(x) + m \bar{z}(x)\right). \end{aligned}$$

Here $\bar{z}(x) = \frac{1}{C}\int_\Omega z(x')\rho(x',x)dx'$ is considered as an average of z in the neighborhood of x by the weight function $\rho(x',x)$. The above equation can be rewritten as

$$E\left[\frac{\partial z(x)}{\partial t}\right] = \frac{C}{Z}\left[(m+1)z(x) + m(\bar{z}(x) - z(x))\right].$$

On the right hand side, the first term is respect to exponential growth tending to generate a scale free distribution of $z(x)$, the second term is a diffusion term which will smooth the distribution of $z(x)$.

3 Between-city transportation

In this section we focus on RIPA on 2-dimensional surface with respect to the case of between-city transportation. First, we consider networks generated by RIPA on the unit square D with periodic boundary conditions. The relevance ρ is given by $f(x) = \exp(-\lambda x)$. In this case the total partition function is:

$$Z = \int_{x \in D} \sum_{j=1}^{N} k_j e^{-\lambda d(x_j,x)} dx$$

Figure 1 represents a special realization of the network. Each circle in the figure represents a city, the center of the circle indicates the locations of the city and the radius indicates the degree, the color(brightness) in the background indicates the logarithm of the local partition function $z(x)$. In Fig.1, we observe a phenomenon that cities tends to gather but big cities tends to separate. Around the greatest city (the capital), we can find bigger city in the area further from the capital. This is because a huge city has two effects: (1) the local partition in its neighbor area is bigger therefore attract more new cities, (2)it will attract more links from new cities therefore inhibit the nearby cities to grow. The second effect is the most significant when we choose small m.

Next, we will investigate the properties of the RIPA network model one by one in this special case, and compare this network model with the BA network and the world airline network (WAN). The later is an empirical network from openfights.org.

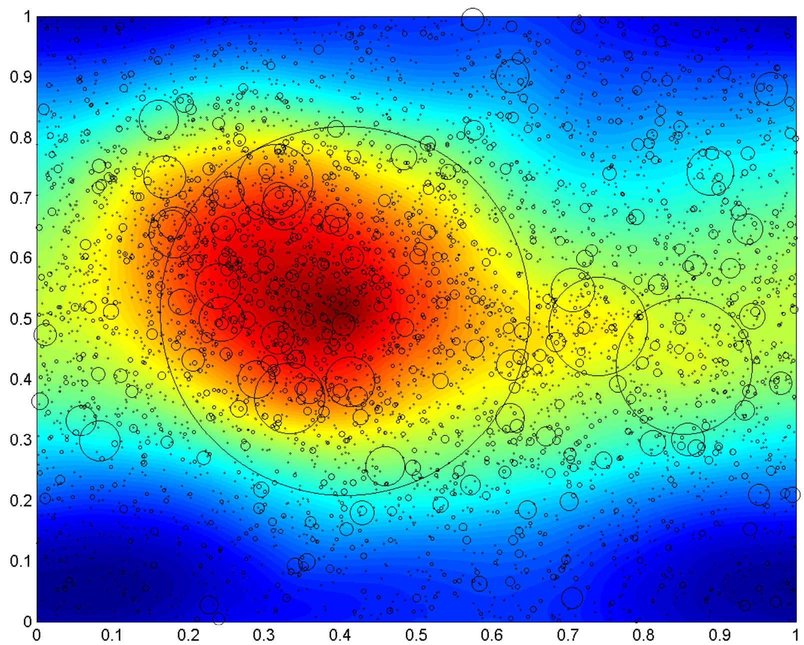
3.1 Degree distribution

Fig. 2 shows that the power-law degree distribution of the RIPA network. As analyzed before, the degree distribution is $N_k \sim k^{-\gamma}$. N_k is the number of vertices with the degree k. The index $\gamma = 3$ as the same as in the BA network model.

3.2 Clustering Coefficient

The clustering coefficient quantifies how well connected are the neighbors of a node in a network. In the RIPA network model, because of the underlying metric space, the "relevance" is naturally transitive, i.e. two objects relevant to the same thing are more likely to be relevant to each other. Consequently, the RIPA network has a significant higher clustering coefficient then the ER or BA networks. Fig.3 shows the clustering coefficients of the RIPA network, the BA network and the WAN network.

Figure 1: Network generated on unit square with periodic boundary condition. $m = 1$, $N = 5000$, $\lambda = 10$. The circles are centered at the locations of the cities and the radii represents their degrees. The background color indicates the logarithm of local partition.



4 Average path-length

In the area of complex networks, we say a network is a "small world" if the average path-length of two arbitrary nodes in the network is no more than the order $O(\ln(N))$ as the network size N grows. There are two different large N limits of this network model. One is the non-extensive limit, for which the metric space keeps the same and the density of nodes increases to infinity. The other is the extensive limit, for which the density of nodes keeps the same and the metric space extends to infinity. In the latter case, an equivalent way is to keep the metric space the same and rescale the metric. For instance, on the unit square, the metric $d(x, y)$ should be rescaled as $d_N(x, y) = \sqrt{N}d(x, y)$, so that the average density of nodes keeps constant as N grows.

According to Fig.4, the RIPA under non-extensive limit is always a small world. The average path-length even lightly decays as N grows. This observation can be interpreted as the transportation in a fixed area becomes more convenient when you have more choices of transition points. We also observe that the RIPA under extensive limit is a small world when the relevance function f has the power-law decay ($f(d) = d^{-2}$), but is not when f has a exponential decay ($f(d) = e^{-\lambda d}$). From the physics aspect, the two relevance functions are analogues of long-range and short-range correlations. So this observation can be concluded as the RIPA network is a small world when the relevance function represents a long-range correlation.

Figure 2: Power-law degree distribution of networks when $m = 1, 5$, $N = 5000, 10000, 20000$, $\lambda = 10$.

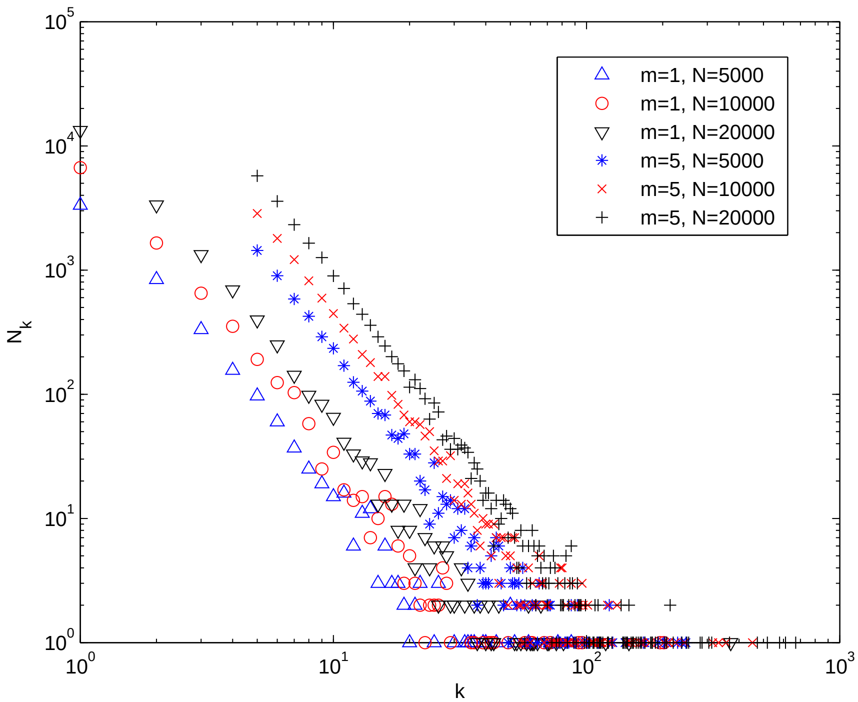


Figure 3: Clustering coefficient C as a function of network size N. RIPA1 for m=3, RIPA2 for m=10.

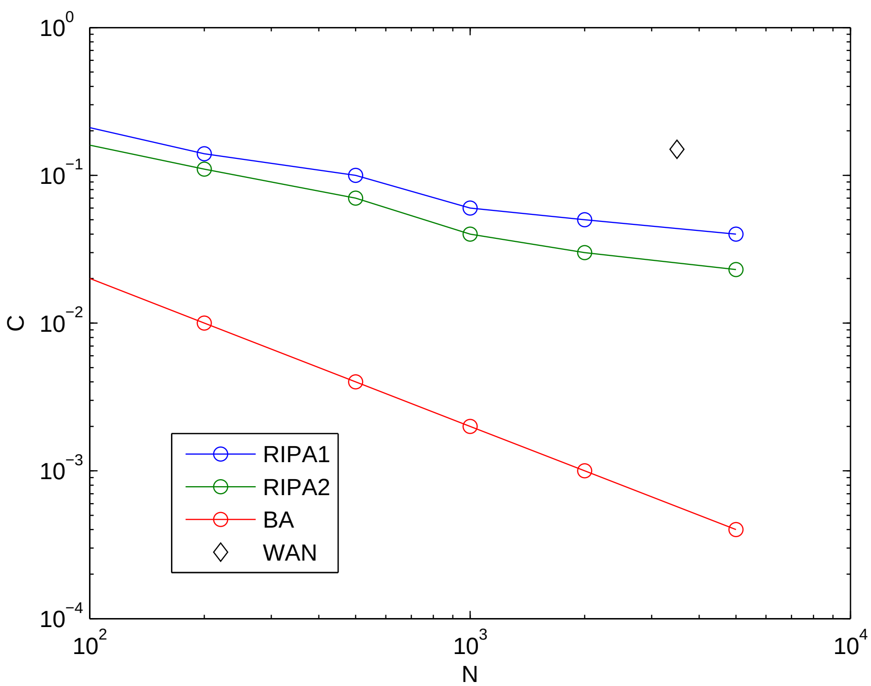
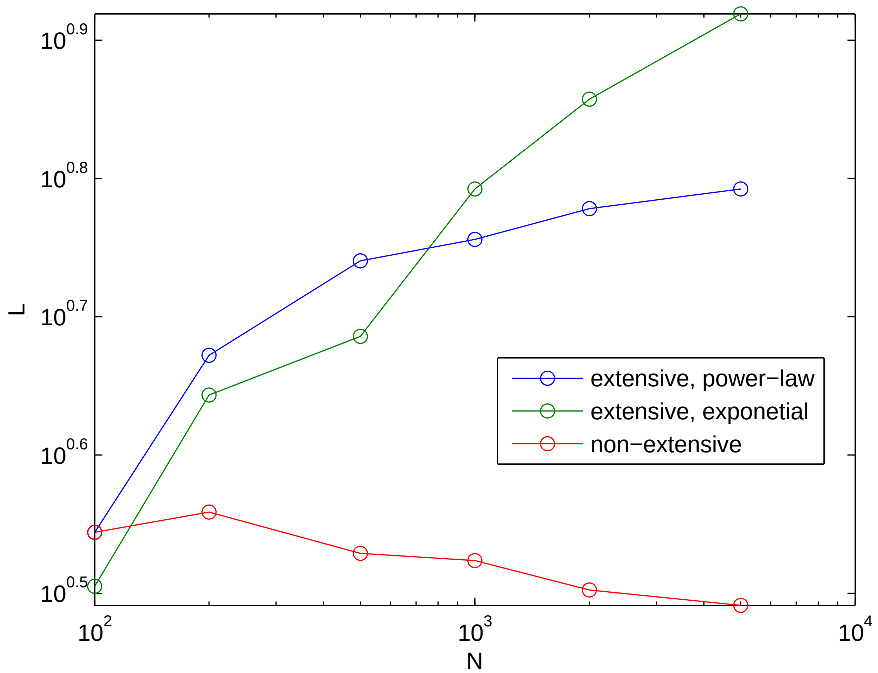


Figure 4: Average path-length L in RIPA network as network size N grows. Red plots are for the RIPA under the non-extensive large N limit. Blue and Green plots are for the RIPA under the extensive large N limit. The blue plot is for the relevance function with power-law decay, the green one is for the relevance function with exponential decay.



The following theorem give a criterion when the RIPA network on two-dimensional space is not a small world.

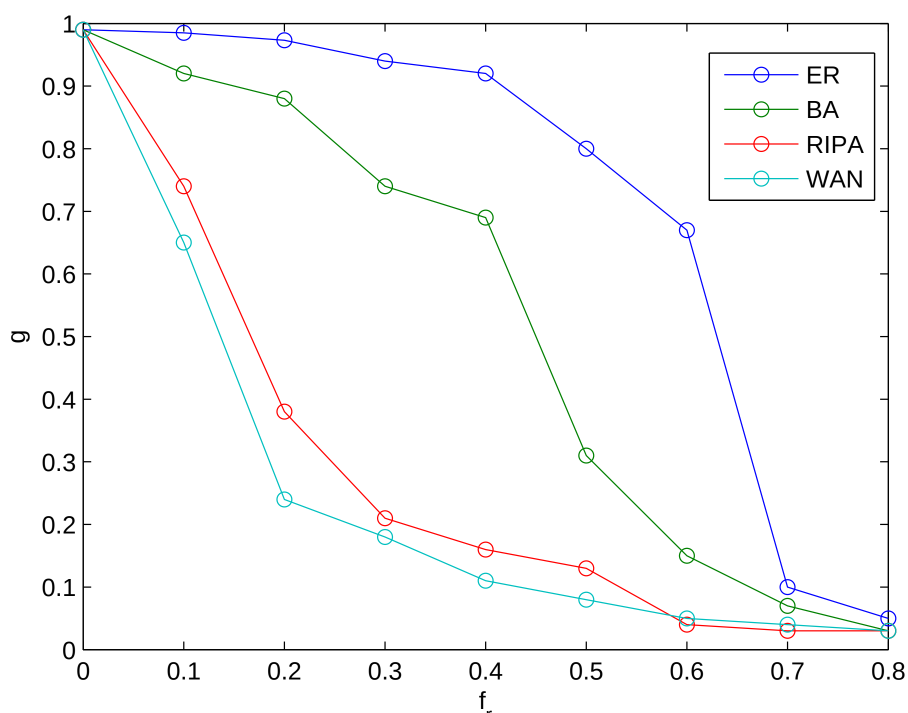
Theorem: The network is not a small world network if the

$$\lim_{r_0 \to \infty} \int_{r=r_0}^{\infty} f(r) r dr < 1$$

5 core-periphery structure

Core-periphery structure is observed in several real world complex networks. In the network with such kind of structure, there is a subnetwork called "core" which is tightly connected, and the complementary subnetwork, the periphery, are fragmental and mostly attached to the core. A significant feature of the core-periphery structure is that the network is vulnerable to the attacks on the core. By successively removing nodes from the core, the whole network will quickly fall into several disconnected parts. The Fig.**??** shows how the giant cluster size decreases as the nodes are removed in the descending order of the degrees. As shown in the figure, the BA network has hubs therefore are more vulnerable to the attacks on the high degree nodes than the ER networks, but it still has a high threshold (about 0.5 in the figure) when the giant cluster size has a fast decay. For RIPA and WAN, however, the giant cluster sizes both decrease quickly at the very beginning. So the RIPA network model captures the core-periphery structure as in the WAN network.

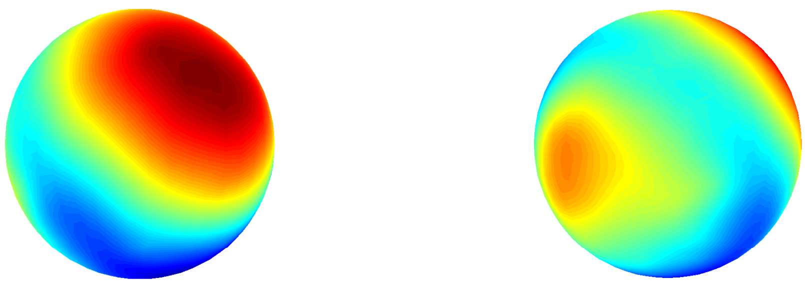
Figure 5: Giant cluster size g after removal f_r fraction of nodes in a descending order of degree.



5.1 RIPA on the Sphere

Similarly, we implement the RIPA on the sphere where the metric is given by spherical distance. As shown in Fig.6, the . Interestingly, some qualitative behavior is quite stable in the simulations, eg. the spherical angle between the first two largest hubs are usually around $0.6\pi - 0.7\pi$. However, this network is not exactly the case of the earth. On the earth, city can only locate on the continents, and the metric is not uniform. The oceans, rivers and mountains may affect the effective distance.

Figure 6: Network generated on sphere with $m = 3$, $N = 5000$, $\lambda = 5$. Two plots are the views of the same sphere from different angles. The color(brightness) indicates the logarithm of local partition.



Acknowledgment

This work was supported in part by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053, by the Army Research Office Grant Nos W911NF-09-1-0254 and W911NF-12-1-0546, and by the Office of Naval Research Grant No. N00014-09-1-0607. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies either expressed or implied of the Army Research Laboratory or the U.S. Government.

References

[1] Erdos, P.; Renyi, A. (1960). "On the evolution of random graphs". Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5: 17-61.

[2] Barabasi, Albert-Laszlo; Albert, Reka (1999). "Emergence of scaling in random networks". Science 286 (5439): 509-512.

[3] Watts, Duncan J.; Strogatz, Steven H. (1998). "Collective dynamics of 'small-world' networks". Nature 393 (6684): 440-442.

[4] Penrose, Mathew: Random Geometric Graphs (Oxford Studies in Probability, 5), 2003.

[5] De Masi, Giulia; et. al (2006). "Fitness model for the Italian interbank money market". Physical Review E 74: 066112.

[6] Dangalchev Ch., Generation models for scale-free networks, Physica A 338, 659 (2004).

[7] J Leskovec, J Kleinberg, C Faloutsos(2005). Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. Proceedings of the eleventh ACM

[8] Alejandro F. Rozenfeld, Reuven Cohen, Daniel ben-Avraham, and Shlomo Havlin(2002). "Scale-Free Networks on Lattices". PRE. VOLUME 89, NUMBER 21.

[9] Tao Zhou Gang Yan and Bing-Hong Wang. Maximal planar networks with large clustering coefficient and power-law degree distribution

[10] Hyperbolic geometry of complex networks Phys. Rev. E 82, 036106 C Published 9 September 2010 Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marin Bogu09

[11] Daniel ben-Avraham, Alejandro F. Rozenfeld, Reuven Cohen, Shlomo Havlin. Geographical Embedding of Scale-Free Networks

[12] Albert-Laszlo Barabasi, Zoltan Dezso, Erzsebet Ravasz, Soon-Hyung Yook and Zoltan Oltvai. Scale-free and hierarchical structures in complex networks

[13] T. Verma, N. A. M. Arujo. and H. J. Herrmann. Revealing the structure of the world airline network.

[14] M. PUCK ROMBACH, MASON A. PORTER, JAMES H. FOWLER, AND PETER J. MUCHA. "CORE-PERIPHERY STRUCTURE IN NETWORKS". SIAM J. APPL. MATH. Vol. 74, No. 1, pp. 167C190.

[15] Paolo Crucitti, Vito Latora, Massimo Marchiori, Andrea Rapisarda (2004). "Error and attack tolerance of complex networks". Physica A: Statistical Mechanics and its Applications. Vol. 340, Issues 1C3, 388-394.